DNA Visual Phasing

Tanner Tolman, AG®

tanner.tolman@familysearch.org
24 April 2022

DNA Visual Phasing

- "Visual Phasing is a process by which the DNA of a set of siblings is assigned to their four grandparents using identified recombination points, without requiring the testing of either the parents or grandparents"-The Visual Phasing Working Group
- Traditionally done by comparing the DNA of three full siblings, but can be done with more or less
- Can be done with half-siblings but only for the common half
- Each person gets exactly half their DNA from mom and dad
- You probably don't have exactly 25% of your DNA from each grandparent
- At any given point you have DNA from only two of your grandparents
- The places where your DNA switches from one grandparent to another are called recombination points

Quality Checks

The Family History Fanatics have been working on a crowd sourcing project to determine how many recombination points there are per chromosome. This is done by comparing known grandparent-grandchild relationships in GedMatch and then counting the number of segments. Their results show that maternal DNA recombines 70% more than paternal DNA on average. For more information on this, check out their excellent video "How often does DNA Recombine" on Youtube

Additionally, the Family History Fanatics have created a "Chromosome Selection Matrix". You can count the number of segments a chromosome has, and it tells you how likely it is to be paternal or maternal. In some cases, it will tell you it has to be a certain one or it may even tell you that the result you found is impossible. For example, their research shows that chromosome 21 can only have three segments if it is maternal and that it will never have four or more. If your visual phasing shows that your DNA breaks these rules, then either your DNA is the first time we have ever observed that or… you messed up your visual phasing. The chromosome selection matrix can be found by going to Youtube and searching for the video "Visual Phasing: A Free Tool to Your Genetic Genealogy Research" by the Family History Fanatics.

Benefits of Visual Phasing

- Makes it easier to figure out how you are related to an unknown cousin
- Reduces the problem of eliminating all potential lines a match can be related on
- It's really fun
- Necessary to accurately reconstruct 75-100% of an ancestor's DNA
- You can make excellent reconstructions at borlandgenetics.com

Reconstructing the DNA of Deceased Ancestors

If you would like to create DNA kits for your deceased ancestors, the best website to do so is borlandgenetics.com. The website has about 20 tools, some that are free and some that require a subscription, that you can use to rebuild an ancestor's by using the DNA of their descendants. The website is user friendly and offers an automated tool called "The Creeper" that can do much of the process for you. You can download your work and upload it into GedMatch as research kits to further help with your research.

Getting Started

- You can test with Ancestry, 23andme, FamilyTreeDNA, MyHeritage, or LivingDNA
- The siblings do not need to have all tested with the same company
- Transfer all their raw data files in GedMatch

Phasing Platform

- You can test with Ancestry, 23andme, FamilyTreeDNA, MyHeritage, or LivingDNA
- The siblings do not need to have all tested with the same company
- Transfer all their raw data files in GedMatch

GedMatch's Chromosome Browser

- When comparing full siblings in GedMatch each chromosome will have segments highlighted with three colors. The three colors are green (full match), yellow (half match), and red (no match). The blue bar underneath simply highlights matches to make them easier to see.
- The places where the siblings inherited DNA from the same grandparents on both sides are green, the places where the siblings inherited from the same grandparent on one side but opposite grandparents on the other side are yellow and the places where they inherited from opposite grandparents on both sides are red.
- Long green segments are usually only seen in full siblings unless there is endogamy. Other relations will only have yellow and red.
- Yellow often has lots of green static in it, and red often has lots of yellow static. Do not worry about that. Worry about the main colors.
- Green sometimes has thin yellow lines through it. Often these are because of incorrectly

reported values along the genome. DNA tests are over 99% accurate but there are about 16 incorrect values per test. If you have several yellow lines next to each other it could actually be a small yellow segment sandwiched between two green segments. On the One-One Autosomal DNA comparison you can see the DNA more clearly by checking the “Full Resolution” box and then scrolling to the part in question. In the image further down in this handout, the green segments have some thin yellow bars. In this case they are of no concern.

Setting It Up Manually in Excel

- Compare siblings A and B, A and C, and B and C in GedMatch’s One-to-one Autosomal DNA comparison.
- Leave the default settings mostly the same, but I recommend checking the box at the bottom that says, “Prevent Hard Breaks (default is to create hard breaks when distance between SNP's exceeds 500,000 base positions):” there are untested regions along the chromosome and checking this box will treat them like one segment
- Each chromosome must be done individually. Take screenshots of one of the chromosomes and put the comparisons in excel. Line up the chromosomes as best you can and try to have them all the same size
- Click and drag the boundaries of the cell columns and line them up with the recombination points. The recombination points are where the chromosomes switch from one color to another.
- Now you need to identify which sibling owned each recombination. This is usually the sibling who switches from one color to another in both of his comparisons. For example, in the image below at the first point siblings A and B go from yellow to green in the same spot that siblings A and C go from yellow to red. Therefore, sibling A probably owns that recombination point.
- Put initials of the siblings or something to help you remember who owns each one above each recombination point.
- Next, I recommend coloring in six bars below the chromosomes with a neutral color. This is where you will fill in the solution. See the image below.
- Finally select four colors. It doesn’t matter what colors you choose put pick two for the top chromosomes for each child and two for the bottom. Be consistent for all three siblings.

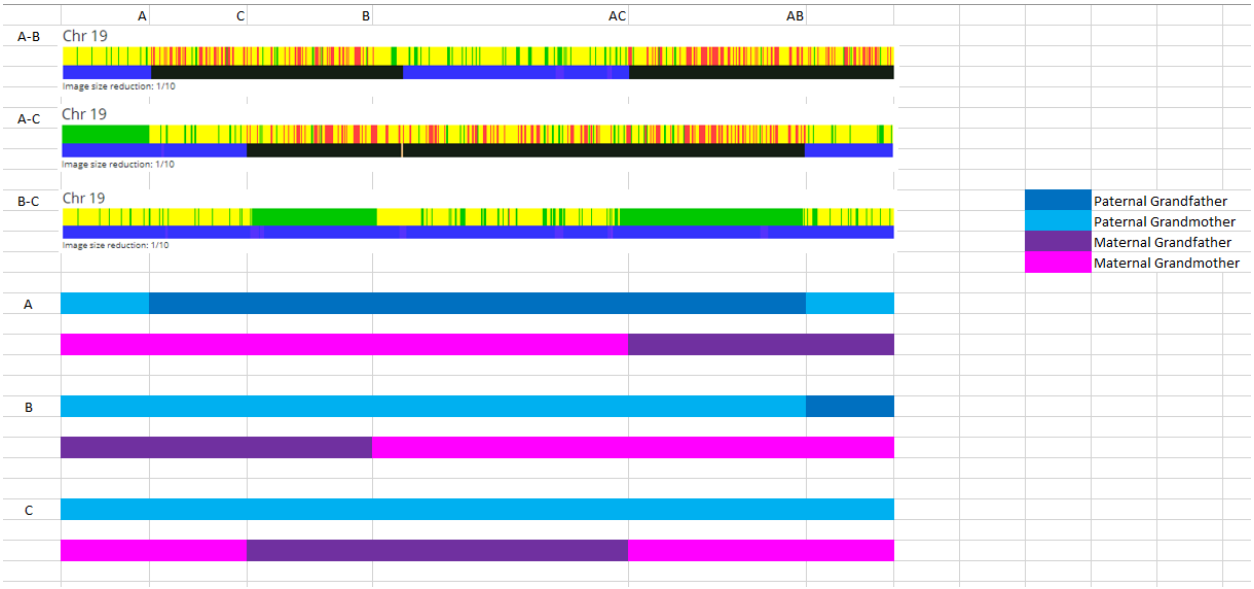
Setting It Up in Steven Fox’s Automated Spreadsheet

- Log in to Facebook and go to the “Visual Phasing Working Group”
- Click on “Files” then click on the “DNA – Visual Phasing Spreadsheet – V2.6” and download it to your computer
- There is also a “Visual Phasing Spreadsheet V2.6” user guide that is a detailed manual
- After having opened the spreadsheet, fill in the names of your four grandparents in the “Grandparent Table”
- Fill out the desired information in the “Sibling Table”
- Filling out the “Cousin Table” is helpful but not necessary to make the program run
- When you have everything the way you want it, click the button that says “build” in the top left corner

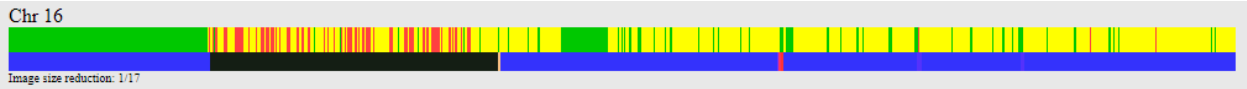
- You will be asked to log into GedMatch with your username and password, do so
- The program will start running automatically, occasionally GedMatch will require a Captcha response to make sure you are a human, if it does so, go to GedMatch scroll down and click the “continue” button
- Do not interfere unless prompted to, overwise you will cause a run-time error and need to start over

Common Problems When Doing Visual Phasing

- Visual phasing is not always straight forward. Sometimes two or more siblings will have a recombination in the same or nearly the same spot or one sibling will have two recombinations in the same spot. Sometimes this will be obvious other times it will be hidden and only come to lot when comparisons against cousins does not match up with the solution you created.
- Here is the solution to an example of two siblings that had a recombination that was not obvious initially:



- At the first point it initially looks like sibling B has a recombination, but the truth is A and C do but not B. At the second point it looks like sibling C has the recombination, but the truth is A and B do but not C.
- Remember that a recombination will usually be for the one that it appears to be for, but it could be that the other two siblings have recombinations in that spot instead
- Comparing against cousins can help prove the places these spots exist or adding a fourth sibling to the mix can help unless the fourth sibling also has a recombination in the same spot which is unlikely.
- Here is a more obvious example of double recombination:



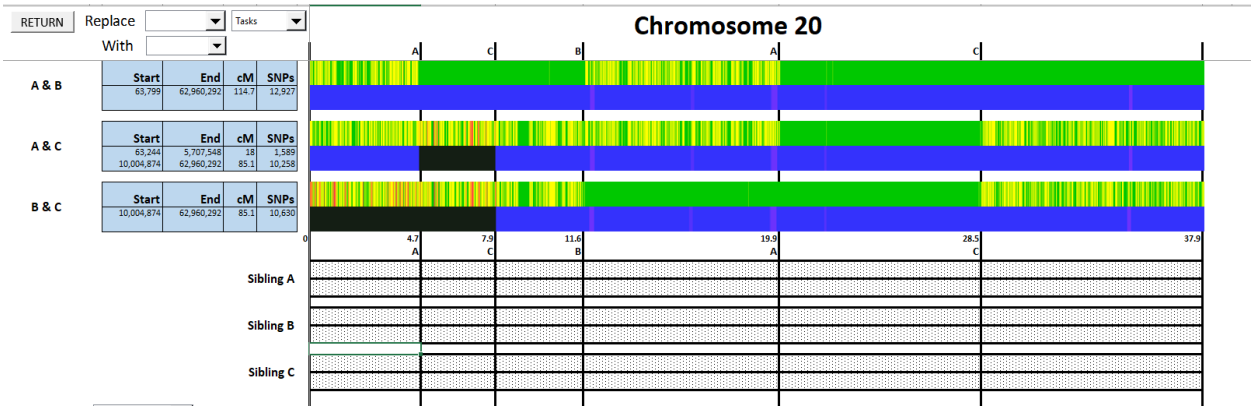
- The only way the siblings could go from green to red is with two recombinations in the same place
- Other times you may need to look at a segment under full resolution to help you determine if a yellow line in a green segment or a red line in a yellow segment is an incorrectly called SNP or a true recombination

Further Information

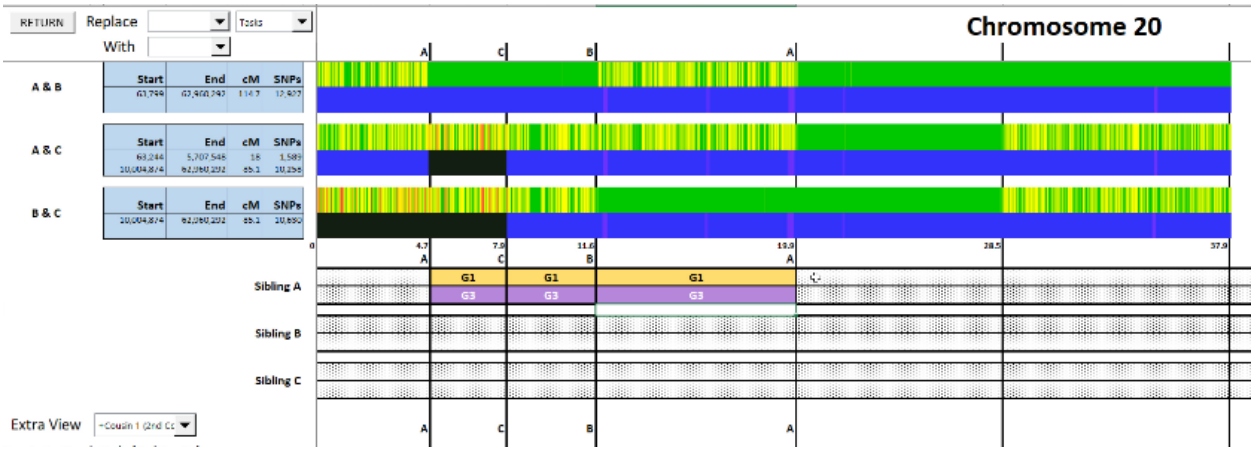
If you would like more information on visual phasing, I strongly recommend the Visual Phasing Working Group on Facebook, the Youtube videos by the Family History Fanatics. There are also others who have made good videos on this topic on Youtube. I also strongly recommend the book *Advanced Genetic Genealogy: Techniques and Case Studies* by Debbie Parker Wayne. That last book has a full chapter completely dedicated to visual phasing. You are also welcome to email me at tanner.tolman@familysearch.org or book an online consultation with me and I would be happy to help you https://go.oncehub.com/ResearchStrategySession.

DNA Visual Phasing: Activity

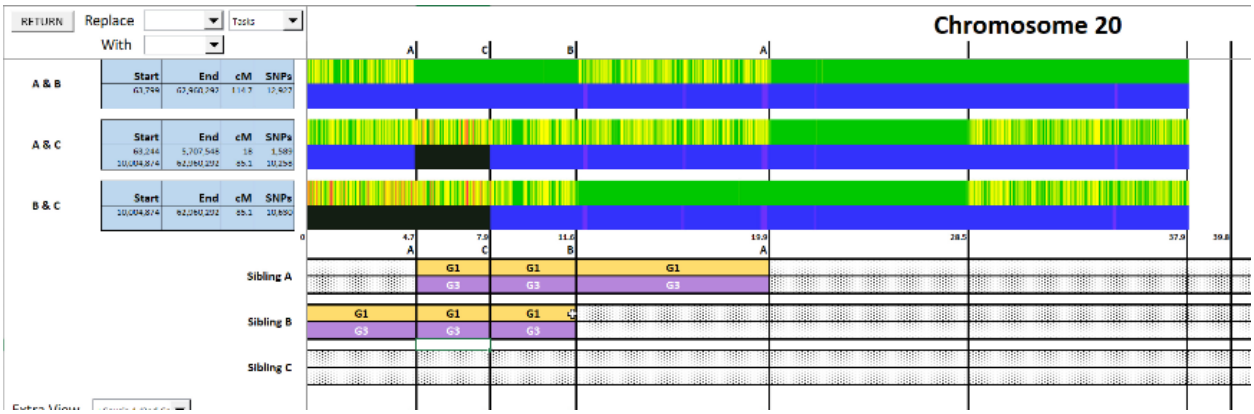
- Let's practice on chromosome 20



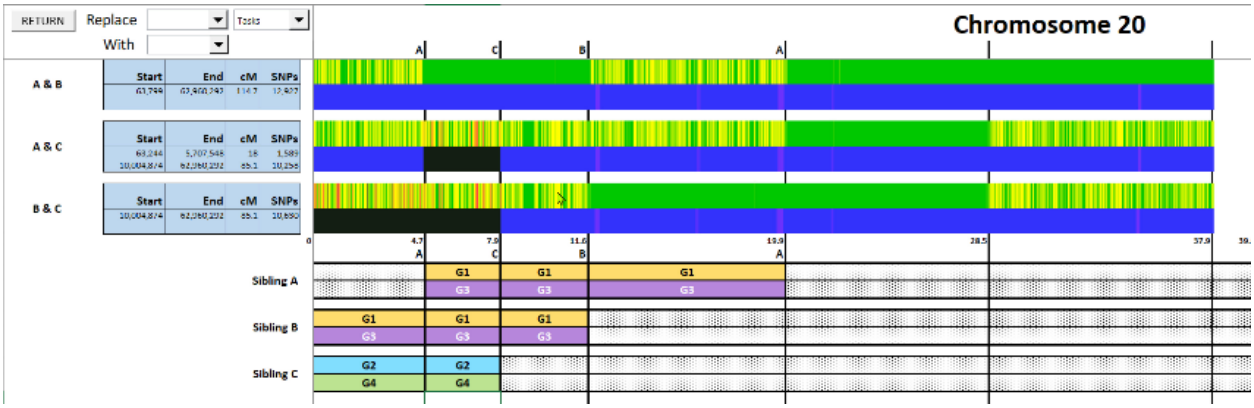
- Pick two colors that you will use on the top copies of each chromosome and two that you will use on the bottom. In this example, peach and blue are on top, green and purple are on bottom.
- Pick a spot on the chromosome for one of the siblings and fill it in with two colors. Just pick a spot and two colors. It doesn't matter and you don't know what is what at this point. Then fill in those same colors to the left and right until you come to a recombination point for that sibling



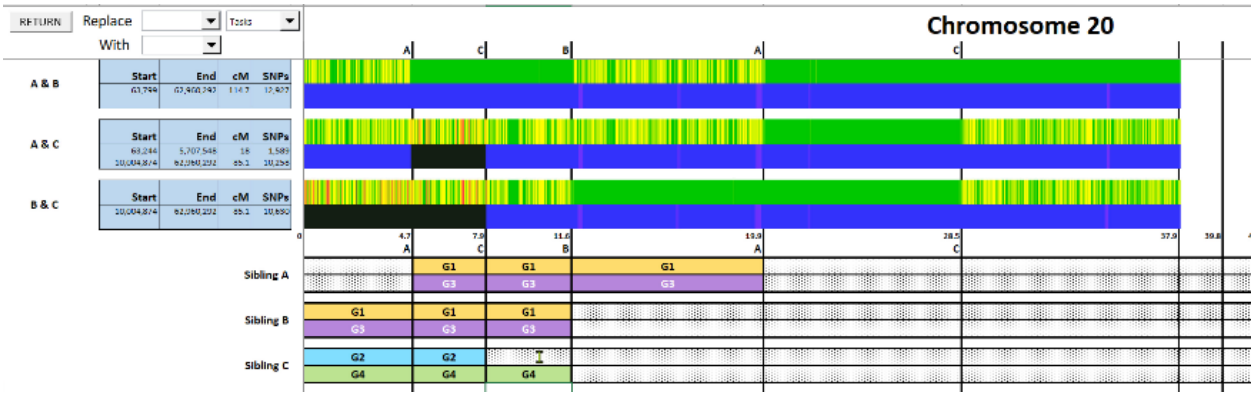
- The spot chosen in this example allows us to fill in some information for siblings B and C as well. In the third column siblings A and B are fully identical. If sibling A is peach and purple, then sibling B must be as well. Then expand that to the left and right until sibling B comes to recombination points



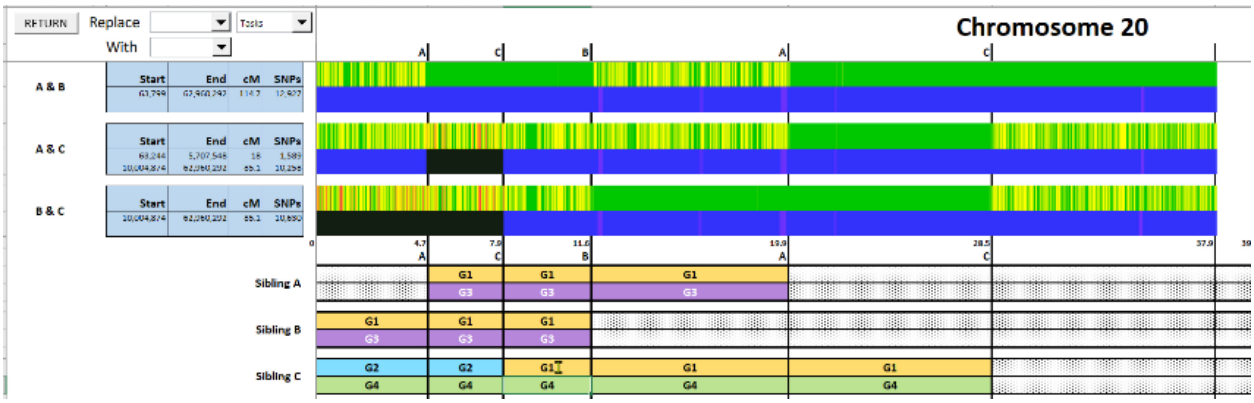
- We can fill in some of sibling C as well. Notice in the 2nd column that sibling C's DNA completely does not match siblings A and B. If siblings A and B are peach and purple, sibling B must be blue and green in that spot, and we can expand that to the left and right to his recombination points too



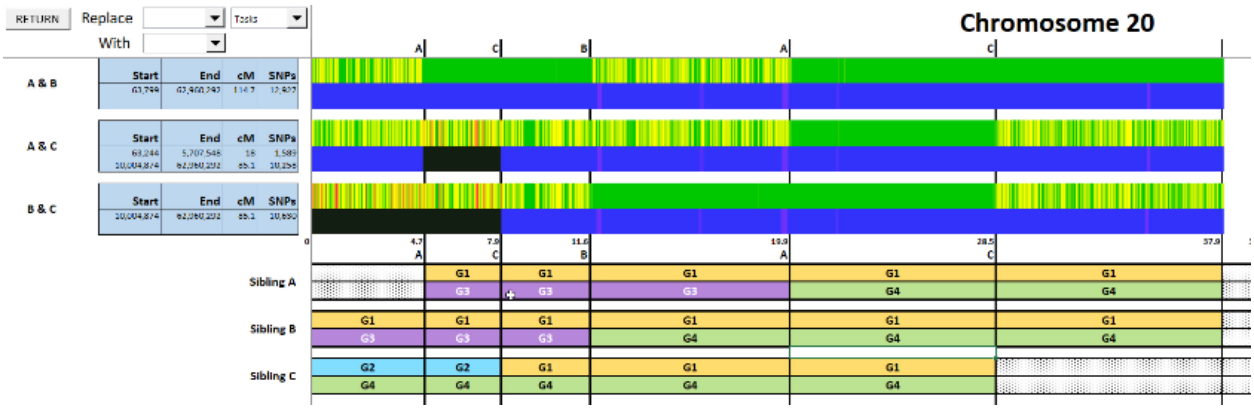
- At this point the colors are still arbitrary meaning any color can be any grandparent. To go further than this point I can pick a spot and fill it with any color but at that point the colors will no longer be arbitrary. Each color will be a specific grandparent even if I don't know which is which. In this example, lets fill in green DNA for sibling C in the bottom column. Why are we picking that place? Just because.



- Filling in that box green has implications that can allow us to fill in most of the chromosome. In that box sibling C is half identical to siblings B and C, if their DNA is different on the bottom chromosomes, then it must be the same on the top chromosomes there. Therefore, if sibling C has green DNA there then they also have peach DNA there. Sibling C's DNA can be filled in as peach and green until he comes to a recombination point.

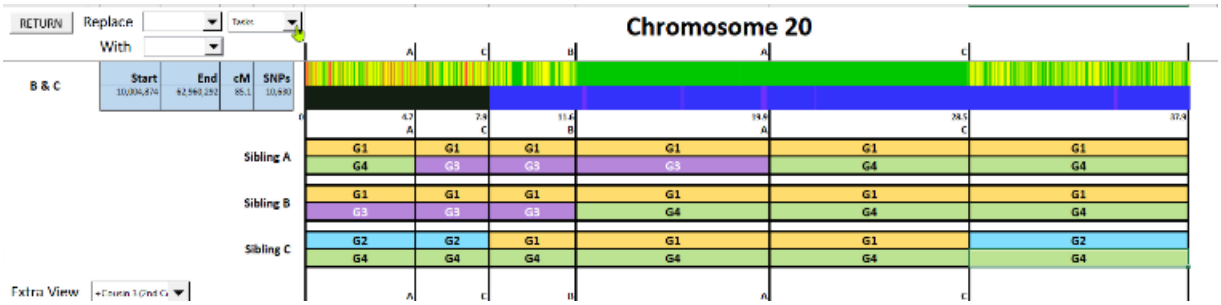


- In the 5th column all three siblings are fully identical, if sibling C has peach and green DNA there then siblings A and B also has peach and green DNA there. Then those colors can be expanded until their recombination points are reached.



- It is now almost completely solved. Only the tips of siblings A and C remain. Since the colors are no longer arbitrary, we cannot just fill in random colors. At this point we need to look at cousins
- This cousin is only related to the maternal grandfather. Their DNA matches siblings A and C on the left side and all three on the right side. On the left side, the segment stops exactly where sibling A has a recombination point, but continues all the way to the third column for sibling C. This must mean that the maternal grandfather is the green color. If the maternal grandfather were the blue, then this cousin would stop matching sibling C right at sibling C's first recombination point.
- Sibling A must have green DNA in the 1st column and also orange DNA since that would make him half identical to sibling C. On the right most side sibling C must have green DNA on bottom and blue DNA on top

- If green is the maternal grandfather, purple must be the maternal grandmother



- The top copies are the paternal grandparents in this case, but we still don't know which is which. This would be determined by comparing against other cousins. In this case, additional cousins revealed that blue was the paternal grandfather and orange was the paternal grandmother

DNA Reconstruction

Tanner Tolman, AG®

tanner.tolman@familysearch.org
24 April 2022

What is Genetic Reconstruction?

To reconstruct the DNA of a deceased ancestor means to create a DNA raw data file for them in word pad or excel file in order to upload it to sites such as GedMatch that looks like what a company would have made for them had they tested personally. It is not reconstructing their physical molecules and has nothing to do with cloning. The results will have little if any use to anyone other than the genealogist.

DNA Raw Data

When you buy a DNA test from a company you are really buying a raw DNA data file. This is a file that records your two values at each tested SNP, for example “AA,” “AG,” “CT” etc. The file can be modified by typing directly into it. If all you really are buying is a word document, then theoretically it would be possible to type such a document for anyone on earth. However, it is statistically impossible to correctly guess all 500K+ values that you would need to make that happen. By carefully analyzing the DNA of enough person’s descendants and close relatives, however, it is possible to create raw data files for a deceased person who is no longer available to test themself.

Borland Genetics

BorlandGenetics.com was debuted at RootsTech London in 2019. It is a website that offers many advanced tools to help you reconstruct the DNA of your ancestors. The website offers both free and subscription versions.

Free Tools to Run on Stereo Kits

- Missing parent: This tool allows you to compare the DNA of a parent and a child. Everywhere they match is deleted out. What remains of the child’s DNA must have been inherited from the missing parent.
- Reverse Phase: This tool allows you to separate out the paternal and maternal DNA of a test taker by comparing their DNA against a child when a parent is not available to test. This tool allows phasing to be done on members of the oldest living generation.
- Phoenix: The DNA of a target person’s descendants is compared against the DNA of cousins and other relatives. Everywhere the children and cousins match must be DNA the children inherited from the target person. This is the tool that is most similar to the Lazarus tool in GedMatch.

- The Dark Side Tool: This tool is the opposite of the Phoenix tool. The DNA of a target person's children is compared against DNA from members of the target person's spouse's family. Everywhere they match but on the children's opposite chromosomes must come from the target parent. For example, you could use your DNA and your mom's sister DNA to reconstruct the DNA of your father.
- Phase Map Locker: More advanced reconstructions will require you to create maps at DNApainter.com. You can download and store those maps here if you need to use them multiple times.
- Humpty Dumpty Merge Option 3: If a living person tests with two or more companies, the results from each can be combined here to create a super kit that has the SNPs from both or all of them
- Standard download: This allows you to download the raw data file you made and upload it to other websites

Free Tools to Run on Mono Kits

- Humpty Dumpty Merge Option 3: Mono to Mono: As you do your reconstruction, you will end up with multiple mono kits for a target person each containing 0-50% of that person's DNA. The mono kits will have no calls in different places and sometimes you need to merge multiple together to get a better mono kit
- Humpty Dumpty Merge Option 2: Mono to Stereo: This lets you merge two mono kits into a stereo kit. For best results you should only do this once and always as the last step in your reconstruction
- Segment Extractor: Sometimes you need to move segments from one ancestor to another such as when you have successfully reconstructed your father's DNA and now you want to start trying to reconstruct his parents a generation back. The segment extractor allows you to overlay a chromosome map created in DNA painter onto a mono kit and move the desired segments to a different ancestor
- Standard download: This allows you to download the raw data file you made and upload it to other websites

Subscription Only Tools

- The Creeper: The creeper is an AI unit that will "creep" through your matches and execute the best possible methods for reconstructing your target person's DNA. The creeper is well built and can do most of the work for you although the creeper may need to ask you questions occasionally such as which side a cousin is related on. One of the coolest things the creeper can do is automated visual phasing. At the end of the process, the creeper will not only create chromosome maps showing which segments the test takers inherited from each of their grandparents, but it will also reconstruct the DNA of both the parents and the grandparents!
- Advanced Download: The advanced download gives allows you to adjust the settings on the raw data file you download. The most important setting is the ability to add white noise into all the gaps of a partial reconstruction

Further Help

- There is a Facebook Group called the “Borland Genetics User Group” you can ask questions there and the members are helpful and friendly
- You are also welcome to email me at tanner.tolman@familysearch.org or book an online consultation with me and I would be happy to help you https://go.oncehub.com/ResearchStrategySession.